SCIENCE CHINA
Life Sciences

SPECIAL ISSUE: Non-coding RNAs
• EDITORIAL •

October 2013 Vol.56 No.10: 865–866
doi: 10.1007/s11427-013-4544-7

Non-coding RNA annotation: Deciphering the second genetic code

QU LiangHu

Biotechnology Research Center, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

Received August 23, 2013; published online August 26, 2013

Citation: Qu L H. Non-coding RNA annotation: Deciphering the second genetic code. Sci China Life Sci, 2013, 56: 865–866, doi: 10.1007/s11427-013-4544-7

"Why do humans have so few genes?" This is one of the 125 most compelling questions that will face scientific inquiry over the next quarter-century, according to the list collected by *Science* in 2005. As we know from the Human Genome Project, less than 2% of the human genome contains protein-coding genes, while the other 98% is composed of non-protein-coding DNA regions. However, what if there are hidden treasures besides DNA elements for replication or transcription? Therefore, the investigation of "the dark genome" was placed in the top 10 "Insights of the decade" that *Science* released in 2010. Scientists have discovered that innumerable non-protein-coding RNA (ncRNA) genes, which were previously thought to be "junk DNA" in the genome, are actually transcribed and may bestow important biological functions.

Recently, the Encyclopedia of DNA Elements Consortium studied 147 cell types in the human body to identify approximately 18400 ncRNA genes covering more than 80% of the human genome. These ncRNAs display special structures and recruit different proteins to form functional complexes throughout cells and even in extracellular fluids. Thus, the human genome is a highly structured RNA machine. However, the current understanding of ncRNAs is only the tip of the iceberg due to enormous difficulties in genome-wide ncRNA gene mining that exist because of the highly divergent and complicated coding rules. Corresponding to the degeneracy, universality and versatility of the genetic code for proteins, the ncRNA code is characteristic of its complexity, diversity and specificity. It compiles codons for sequence motifs that have potential post-transcriptional structural and interactional activity. Comprehensively annotating the ncRNA genes, that is, deciphering "the second genetic code" from eukaryotes to prokaryotes is the major challenge for life sciences in the 21st century.

The rapid development of next generation sequencing technologies has greatly advanced genome decoding. The production of sequencing data is growing at an exponential rate that outstrips Moore's Law. These data will lead us into a "biological big data" era and provide unprecedented opportunities to reshape the transcriptomic research landscape. "The biology of ncRNA genes" as a new gene science has come of age and is composed of three main areas:

(i) ncRNA informatics. Establish and develop an RNomics platform that integrates supercomputers and high-throughput sequencing technology to obtain, collect, and process the massive amount of bioinformation being generated, and decipher the structure and function of genomes from both protein-coding and non-protein-coding gene annotations, finally demonstrating a modern RNA world.

(ii) ncRNA biology. Aim at the core physiological and genetic issues by revealing ncRNAs as key regulators in a synergistic network together with proteins involving cell function and fate determination, and parsing the molecular and cellular mechanisms of ncRNA origin and evolution in the processes of genetics, epigenetics and acquired genetics.

(iii) ncRNA gene resources and technologies. Mine novel gene resources to improve production in both agriculture and livestock and develop ncRNA-based translational medicine, including using RNA intervention and the RNA index for health and disease applications.

email: lssqlh@mail.sysu.edu.cn

In June 2012, Chinese scientists held the 426 Xiangshan Conference on "The function and mechanisms of ncRNAs in the major biological processes," which reviewed ncRNA research in China and prospects for future research directions. To introduce the current achievements and progress in ncRNA research, *Science China Life Sciences* has organized this special issue on ncRNAs in the hope that this publication will promote collaborative innovation in ncRNA research with multiple disciplines, including informatics, biology and medicine.

Biographical Sketch

Dr. Qu LiangHu graduated from Wuhan University in 1982 and obtained his Ph.D. degree (Docteur d'Etat) from Paul Sabatier University, France in 1986. He received "Distinguished Young Scholar" from the National Natural Science Foundation in 1995, and was honored "the Cheung Kong Scholar" by the Ministry of Education of China in 1999. From 2000 to 2010, he severed as the Director of the Key Laboratory of Gene Engineering of the Ministry of Education of China. He currently is the Professor and Director of Biotechnology Research Center, Sun Yat-sen University, and the chief scientist of the National Basic Research (973) Program of China on the regulatory mechanism of human non-coding RNAs in cell function and fate determination. He has published more than 140 articles in RNA research field.